

The evolving beta coalescent

Götz Kersting*, Jason Schweinsberg† and Anton Wakolbinger*

February 20, 2014

Abstract

In mathematical population genetics, it is well known that one can represent the genealogy of a population by a tree, which indicates how the ancestral lines of individuals in the population coalesce as they are traced back in time. As the population evolves over time, the tree that represents the genealogy of the population also changes, leading to a tree-valued stochastic process known as the evolving coalescent. Here we will consider the evolving coalescent for populations whose genealogy can be described by a beta coalescent, which is known to give the genealogy of populations with very large family sizes. We show that as the size of the population tends to infinity, the evolution of certain functionals of the beta coalescent, such as the total number of mergers, the total branch length, and the total length of external branches, converges to a stationary stable process. Our methods also lead to new proofs of known asymptotic results for certain functionals of the non-evolving beta coalescent.

Keywords: beta coalescent, evolving coalescent, total branch length, total external length, number of mergers, stable moving average processes.

AMS MSC 2010: Primary 60K35, Secondary 60F17, 60G52, 60G55, 92D15.

1 Introduction

In 1999, Pitman [14] and Sagitov [15] introduced coalescents with multiple mergers, also known as Λ -coalescents. These processes are continuous-time Markov processes taking values in the set of partitions of \mathbb{N} , and as time goes forward, blocks of the partition merge together. For any finite measure Λ on $[0, 1]$, the Λ -coalescent is defined by the property that whenever the restriction of the process to $\{1, \dots, n\}$ has b blocks, each possible transition that involves b blocks merging into one happens at rate

$$\lambda_{b,k} = \int_0^1 p^{k-2} (1-p)^{b-k} \Lambda(dp),$$

and these are the only transitions that occur. This means that when there are b blocks, the total rate of all mergers is

$$\lambda_b = \sum_{k=2}^b \binom{b}{k} \lambda_{b,k}.$$

When Λ is a unit mass at zero, only two blocks ever merge at a time, and each transition that involves two blocks merging into one happens at rate 1, so we get the celebrated Kingman's

*Supported in part by DFG SPP 1590 *Probabilistic Structures in Evolution*

†Supported in part by NSF Grant DMS-1206195

coalescent [10]. When Λ is the uniform distribution on $[0, 1]$, the Λ -coalescent is known as the Bolthausen-Sznitman coalescent [4].

Coalescent processes arise naturally in population genetics, where they are used to model the genealogy of populations. The genealogy of a population of size n can be modeled by a process $(\Pi_n(r), r \geq 0)$ taking its values in the set of partitions of $\{1, \dots, n\}$. The integers i and j are in the same block of the partition $\Pi_n(r)$ if and only if the i th and j th individuals in the population have the same ancestor r units back in time. Typically, Kingman's coalescent is used to model the genealogy of populations. However, in some circumstances, other Λ -coalescents could describe the genealogy of a population. For example, it was shown in [17] that if the probability of an individual having k or more offspring is proportional to $k^{-\alpha}$, where $1 < \alpha < 2$, then the genealogy of the population is best described by the Λ -coalescent in which Λ is the Beta($2 - \alpha, \alpha$) distribution. This process will hereafter be called the Beta($2 - \alpha, \alpha$)-coalescent, or simply the beta coalescent. Beta coalescents are thus natural models for populations with large family sizes, and predictions from beta coalescents were shown in [19] to fit genetic data from some marine species.

There has been recent interest in describing not only the genealogy of a population at a fixed time but also how the genealogy of a population changes over time. At any given time, the genealogical structure described by the coalescent process can also be represented by a tree. The shape of this tree changes over time, and the associated tree-valued process $(\mathcal{T}_n(t), t \in \mathbb{R})$ is known as an evolving coalescent. For populations whose genealogy at a fixed time is described by Kingman's coalescent, the associated evolving coalescent was studied in [12, 13, 8]. The evolving Bolthausen-Sznitman coalescent was studied in [18].

In the present paper, we study the evolving coalescent for populations whose genealogy is given by the Beta($2 - \alpha, \alpha$)-coalescent, where $1 < \alpha < 2$. To this end we consider a time rescaling depending on n , the *scaled* time being $s = n^{\alpha-1}t$. On the original time scale, merging events occur at a rate of order n^α ; see formula (23) below. Therefore the number of coalescent events in one unit of the scaled time is of order n , meaning that the rate at which a specific one of the n lineages takes part in a merging event is of order 1. Thus, while the original time t captures *evolutionary time*, the scaled time s plays the role of a *generation time*.

We show that as $n \rightarrow \infty$, the distribution of certain functionals of the beta coalescent converges to a stable distribution of index α . For the evolving beta coalescent, the distribution of these functionals evaluated at times $s_1 < s_2 < \dots < s_d$ converges as $n \rightarrow \infty$ to a multivariable stable distribution of index α . Examples of functionals that fit into this framework include the total number of merger events before all of the lineages have coalesced into a single lineage, the total length of all branches in the tree, and the total length of all external branches in the tree.

As a typical result, which follows by combining Example 7 and Corollary 5 below, we state the following theorem, which shows how the total branch length of the coalescent tree evolves over time.

Theorem 1. *Let $\mathcal{L}_n(t)$ be the total branch length of $\mathcal{T}_n(t)$, $t \in \mathbb{R}$. Then for $1 < \alpha < \frac{1}{2}(1 + \sqrt{5})$ the sequence of processes*

$$n^{\alpha-1-1/\alpha} \left(\mathcal{L}_n(n^{1-\alpha}s) - \frac{\alpha(\alpha-1)\Gamma(\alpha)n^{2-\alpha}}{2-\alpha} \right), \quad -\infty < s < \infty,$$

converges in finite-dimensional distributions as $n \rightarrow \infty$ to the moving average process

$$\int_0^\infty g(r) dL_{s-r}, \quad -\infty < s < \infty, \quad (1)$$

where $g(r) = (\alpha - 1)(\alpha\Gamma(\alpha))^{\frac{1}{\alpha-1}}(r + \alpha\Gamma(\alpha))^{-\frac{2-\alpha}{\alpha-1}}$ and $(L_s)_{-\infty < s < \infty}$, is a mean zero Lévy process with $L_0 = 0$ and Lévy measure

$$\frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)}u^{-1-\alpha}du, \quad u > 0. \quad (2)$$

The time reversal in the integrator is explained in more detail in Section 2.5. The stochastic integral is well-defined for $\alpha < \frac{1}{2}(1 + \sqrt{5})$. Above this threshold, the statement fails to be true even for the fixed value $s = 0$, as shown in [9]. Note that the function $g(r)$ decreases as a power of r , and consequently the moving average process defined in (1) is not Markovian. The exponent $-(2 - \alpha)/(\alpha - 1)$ tends to $-\infty$ as $\alpha \rightarrow 1$. In the limiting case $\alpha = 1$, which corresponds to the Bolthausen-Sznitman coalescent, it is known that the evolution of the total branch length converges to a moving average process defined as in (1), but with $g(r) = e^{-r}$. In this case the limit process is Markovian. For details, see [18].

Our approach consists of deriving asymptotic expansions for suitable functionals by means of Poisson integrals. Thereby we rediscover some known results for functionals of the static (non-evolving) beta coalescent, e.g. for its total length \mathcal{L}_n and total external length ℓ_n . Moreover by means of the Poisson integral representations we get hold also on their (properly scaled) joint distributions, which are asymptotically multivariate stable. This allows us to calculate, for example, the asymptotic distribution of ℓ_n/\mathcal{L}_n (see Example 9 in Section 2.6).

We give a precise construction of the evolving beta coalescent, as well as a precise statement of the main results and a few examples in section 2. Proofs are given in section 3.

2 Framework, main results, and examples

2.1 Construction of the evolving beta coalescent

We give here a precise construction of the evolving Λ -coalescent, which is modeled after the Poisson process construction of the Λ -coalescent given in [14] and is similar to the construction of the evolving Bolthausen-Sznitman coalescent in [18]. Let Λ be a finite measure on $(0, 1]$. We will construct a population of fixed size n defined for all times $t \in \mathbb{R}$ whose genealogy is given by the Λ -coalescent. Individuals in the population will be labeled $1, \dots, n$.

Let $\Upsilon = \Upsilon_n$ be a Poisson point process on $\mathbb{R} \times (0, 1] \times [0, 1]^n$ with intensity measure

$$dt \times p^{-2}\Lambda(dp) \times dv_1 \dots dv_n.$$

Suppose (t, p, v_1, \dots, v_n) is a point of Υ . If zero or one of the points v_1, \dots, v_n is less than p , then no change in the population occurs at time t . However, if $k \geq 2$ of these points are less than p , so that $v_{i_1} < \dots < v_{i_k} < p$, then at time t , the individuals labeled i_2, \dots, i_k all die, and the individual labeled i_1 gives birth to $k - 1$ new individuals who take over the labels i_2, \dots, i_k . This implies that if we are following the genealogy of the population backwards in time, the lineages labeled i_1, \dots, i_k will all coalesce at time t . To see that the Λ -coalescent describes the genealogy of this population, note that the rate of events that cause the lineages i_1, \dots, i_k to coalesce is

$$\int_0^1 p^k(1-p)^{n-k} \cdot p^{-2}\Lambda(dp) = \lambda_{n,k}.$$

This construction is well-defined because the rate of changes in the population is bounded above by

$$\int_0^1 \binom{n}{2} p^2 \cdot p^{-2}\Lambda(dp) < \infty.$$

Note that, since we are ordering the genealogy with respect to the v_i , and not in a lookdown manner with respect to the indices i , we do not have strong consistency in n .

Although this construction works whenever $\Lambda(\{0\}) = 0$, we will hereafter restrict ourselves to the case in which Λ is the $\text{Beta}(2 - \alpha, \alpha)$ distribution. For each $t \in \mathbb{R}$, there will be a different realization of the beta coalescent which describes the genealogy of the population at time t . We denote the corresponding coalescent tree (read off from the genealogy backwards from time t) by $\mathcal{T}_n(t)$ and call the process $(\mathcal{T}_n(t), t \in \mathbb{R})$ the evolving beta coalescent.

2.2 Two Poisson processes

The evolving beta coalescent is constructed from a Poisson process Υ_n on $\mathbb{R} \times (0, 1] \times [0, 1]^n$ with intensity measure

$$dt \times \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} p^{-1-\alpha} (1-p)^{\alpha-1} dp \times dv_1 \dots dv_n.$$

In this section, we will construct two other Poisson processes, denoted by Ψ_n and Θ_n , that will be useful for analyzing the evolving beta coalescent and the static (non-evolving) beta coalescent back from time 0, respectively.

First, we obtain a Poisson process Υ' on $\mathbb{R} \times \mathbb{R}^+$ in two steps by discarding all but the first two coordinates of the points of Υ_n , and then augmenting these points with the points of an independent Poisson process with intensity

$$dt \times \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} p^{-1-\alpha} (1-q(p)) dp,$$

where $q(p) = 0$ if $p \geq 1$ and $q(p) = (1-p)^{\alpha-1}$ if $0 < p < 1$. Note that Υ' has intensity

$$dt \times \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} p^{-1-\alpha} dp. \quad (3)$$

From Υ' , we can then obtain a Poisson process Ψ_n via the mapping

$$(t, p) \mapsto (s, u) := (n^{\alpha-1}t, n^{1-1/\alpha}p).$$

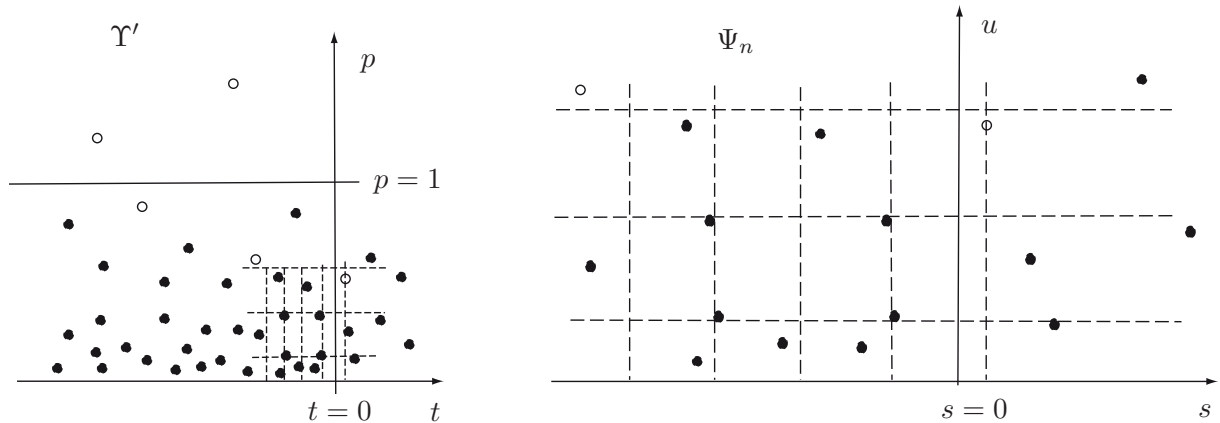


Figure 1: The process Υ' contains the points (marked by \bullet) from the first two coordinates of the points of Υ , plus additional points (marked by \circ) that make up for the difference between the intensities $p^{-1-\alpha}(1-p)^{\alpha-1}1_{p \leq 1}$ and $p^{-1-\alpha}$, $p > 0$. The point process Ψ_n arises from Υ' through the transformation $(t, p) \mapsto (s, u) = (n^{\alpha-1}t, n^{1-1/\alpha}p)$.

That is, if (t, p) is a point of Υ' , then $(n^{\alpha-1}t, n^{1-1/\alpha}p)$ is a point of Ψ_n . It is straightforward to verify that the intensity of Ψ_n is also given by (3), now with (t, p) replaced by (s, u) .

The reason for considering the rescaled Poisson process Ψ_n is that the fluctuations in the behavior of the beta coalescent that will be important for studying the evolving coalescent are those that happen after the coalescent has evolved for a time which is $O(n^{1-\alpha})$. Also, the largest mergers on this time scale affect approximately a fraction $O(n^{-1+1/\alpha})$ of the blocks, in other words a number of $O(n^{1/\alpha})$ blocks out of the n initial blocks. If (s, u) is a point of Ψ_n , then if this point corresponds to a point of Υ_n (i.e. unless it corresponds to a point of Υ' that does not belong to Υ_n), at time $n^{1-\alpha}s$ there is an event during which approximately a fraction $n^{-1+1/\alpha}u$ of the blocks merge together.

We now define another Poisson process Θ_n on $(0, 1] \times \mathbb{R}^+$ which is useful for studying the static beta coalescent that describes the genealogy of the population at time 0. Writing

$$r = -s$$

for the *reverse* time, which is the coalescence time direction, we obtain Θ_n by first restricting Ψ_n to all points (s, u) with $s \leq 0$ and then applying the mapping $(s, u) \mapsto (x, y)$, $t \leq 0$, to the remaining points, where

$$x = m(r), \quad y = m(r)u, \quad (4)$$

for

$$m(r) = \left(\frac{\alpha\Gamma(\alpha)}{r + \alpha\Gamma(\alpha)} \right)^{1/(\alpha-1)}, \quad r \geq 0. \quad (5)$$

This quantity is the asymptotic proportion of the number of blocks that have not yet coalesced by the reverse time r ; see Lemma 13 below.

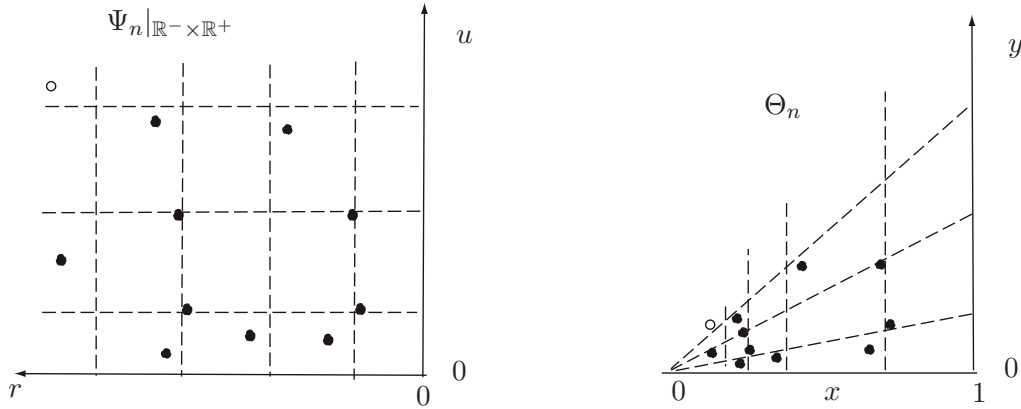


Figure 2: The point process Θ_n arises from $\Psi_n|_{\mathbb{R}^- \times \mathbb{R}^+}$ through the transformation $(-r, u) \mapsto (x, y) = (m(r), m(r)u)$. Here, $r = -s$, and $m(r)$ is given by formula (5).

To calculate the intensity of Θ_n , we invert the mapping given by (4) to get

$$r = \alpha\Gamma(\alpha) \left(\frac{1}{x^{\alpha-1}} - 1 \right), \quad u = \frac{y}{x}.$$

It follows that

$$\frac{ds}{dx} = -\frac{dr}{dx} = \alpha(\alpha-1)\Gamma(\alpha)x^{-\alpha}, \quad \frac{du}{dy} = \frac{1}{x}.$$

Since $\frac{ds}{dy} = 0$, the Jacobian determinant of $(x, y) \mapsto (s, u)$ is $D(x, y) = \alpha(\alpha-1)\Gamma(\alpha)x^{-\alpha-1}$ and the intensity of Θ_n is

$$\nu(dx, dy) = \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} \left(\frac{y}{x}\right)^{-1-\alpha} D(x, y) dx dy = dx \times \frac{\alpha(\alpha-1)}{\Gamma(2-\alpha)} y^{-\alpha-1} dy, \quad (6)$$

which again does not depend on n .

Suppose (x, y) is a point of Θ_n . Because the number of blocks in a beta coalescent at time r is approximately $m(r)n$, the merger corresponding to (x, y) occurs when approximately a fraction xn blocks remain, and the number of blocks that merge is approximately $n^{1/\alpha-1} \cdot yn = n^{1/\alpha}y$. Therefore, whereas the second coordinate of Ψ_n approximates the *fraction* of blocks lost due to a merger, the second coordinate of Θ_n represents the *number* of blocks lost due to a merger. The fact that the Poisson process Θ_n is homogeneous in time reflects the fact that the distribution of the number of blocks lost in the first merger tends to a limit as the number of blocks at time zero tends to infinity; see section 3.1.

2.3 Review of results on stable laws and Poisson integrals

To state and prove our main results, we will need to review some results on both univariate and multivariate stable distributions. We will restrict ourselves to the stable distributions of index α , where $1 < \alpha < 2$. Following the notation of [16], we write in the univariate case $Z \sim S_\alpha(\sigma, \beta, \mu)$ if the characteristic function of the random variable Z is given by

$$E(e^{i\theta Z}) = \exp \left(i\theta\mu - \sigma^\alpha |\theta|^\alpha \left(1 - i\beta \operatorname{sgn}(\theta) \tan \left(\frac{\pi\alpha}{2} \right) \right) \right),$$

with location parameter μ (which equals $E(Z)$ for $1 < \alpha < 2$), scale parameter $\sigma > 0$, and skewness parameter $\beta \in [-1, 1]$. Here $\operatorname{sgn}(\theta) = 1$ if $\theta > 0$, $\operatorname{sgn}(\theta) = -1$ if $\theta < 0$, and $\operatorname{sgn}(0) = 0$. We only deal with the case $\mu = 0$.

When $\mu = 0$, $\beta = 1$, the characteristic function of Z can also be written in the Lévy-Khinchine form

$$E(e^{i\theta Z}) = \exp \left(b \int_0^\infty (e^{i\theta y} - 1 - i\theta y) y^{-1-\alpha} dy \right)$$

(since $\alpha > 1$, we may and do avoid here the customary truncation within the integrand), where

$$\sigma^\alpha = \frac{b\pi}{2 \sin(\frac{\pi\alpha}{2}) \Gamma(\alpha+1)}. \quad (7)$$

One can then construct Z in the following way. Consider a Poisson point process $\sum_{i \geq 1} \delta_{y_i}$ on $(0, \infty)$ with intensity $by^{-1-\alpha} dy$, and for $\varepsilon > 0$ define

$$S(\varepsilon) := \sum_{y_i \geq \varepsilon} y_i, \quad Z(\varepsilon) = S(\varepsilon) - \frac{b\varepsilon^{1-\alpha}}{\alpha-1}.$$

Then the limit

$$Z = \lim_{\varepsilon \downarrow 0} Z(\varepsilon) \text{ a.s.} \quad (8)$$

exists and obeys

$$E(Z) = E(Z(\varepsilon)) = 0, \quad \text{Var}(Z - Z(\varepsilon)) = \int_0^\varepsilon y^2 \cdot by^{-1-\alpha} dy = \frac{b\varepsilon^{2-\alpha}}{2-\alpha}.$$

To facilitate comparisons with results in [5, 9], we note that if $c > 0$ and if Z has a stable distribution such that

$$E(Z) = 0, \quad P(Z > z) = z^{-\alpha}(1 + o(1)), \quad P(Z < -z) = o(z^{-\alpha}) \quad (9)$$

as $z \rightarrow \infty$, then $cZ \sim S_\alpha(\sigma, 1, 0)$ with σ given by (7) with $b = \alpha c^\alpha$, which means

$$\sigma = c \left(\frac{\pi}{2 \sin(\frac{\pi\alpha}{2}) \Gamma(\alpha)} \right)^{1/\alpha}. \quad (10)$$

Note that the results in [5, 9] are stated with $-Z$ in place of Z .

Next we define the notion of an α -stable random measure, following sections 3.3 and 3.12 of [16], again only for $\beta = 1$. Let (E, \mathcal{E}, ρ) be a measure space, and let $\mathcal{E}_0 = \{A \in \mathcal{E} : \rho(A) < \infty\}$. Then an α -stable random measure with control measure ρ is defined as a countably additive function M which assigns a random variable to each set $A \in \mathcal{E}_0$, satisfying the following properties:

1. If $A \in \mathcal{E}_0$, then $M(A) \sim S_\alpha(\rho(A)^{1/\alpha}, 1, 0)$.
2. If A_1, \dots, A_k are disjoint sets in \mathcal{E}_0 , then $M(A_1), \dots, M(A_k)$ are independent.

To construct M , consider a Poisson point process Ξ on $E \times \mathbb{R}^+$ with intensity $a\rho(dx) \times y^{-1-\alpha} dy$, where

$$a = \frac{2 \sin(\frac{\pi\alpha}{2}) \Gamma(\alpha + 1)}{\pi}. \quad (11)$$

Then the second coordinates of those points of Ξ that fall in $A \times \mathbb{R}^+$ form a Poisson point process on \mathbb{R}^+ with intensity $a\rho(A)y^{-1-\alpha} dy$. If one constructs $M(A)$ just as Z is obtained in (8) with $b = a\rho(A)$, then $M = M^\Xi$ is an α -stable random measure with control measure ρ (and $\beta = 1$) obtained from Ξ .

As noted in section 3.4 of [16], if $f : E \rightarrow \mathbb{R}$ is a function such that $\int_E |f(x)|^\alpha \rho(dx) < \infty$, then one can define the integral

$$I(f) = \int_E f(x) M(dx) \quad (12)$$

by approximating f with simple functions. It is shown in [16] that $I(f) \sim S_\alpha(\sigma_f, \beta_f, 0)$, where

$$\sigma_f = \left(\int_E |f(x)|^\alpha \rho(dx) \right)^{1/\alpha}, \quad \beta_f = \frac{\int_E |f(x)|^\alpha \text{sgn}(f(x)) \rho(dx)}{\int_E |f(x)|^\alpha \rho(dx)}. \quad (13)$$

In particular it follows for any linear combination

$$\theta_1 I(f_1) + \dots + \theta_d I(f_d) = I(\theta_1 f_1 + \dots + \theta_d f_d) \sim S(\sigma_f, \beta_f, 0)$$

with $f = \theta_1 f_1 + \dots + \theta_d f_d$. In the case $\alpha > 1$ this implies that the joint distribution of $I(f_1), \dots, I(f_d)$ is a multivariate α -stable distribution. The characteristic function is

$$\begin{aligned} & E(e^{i(\theta_1 I(f_1) + \dots + \theta_d I(f_d))}) \\ &= \exp \left(- \int_E |\theta_1 f_1 + \dots + \theta_d f_d|^\alpha (1 - i \operatorname{sgn}(\theta_1 f_1 + \dots + \theta_d f_d) \tan(\frac{\pi\alpha}{2})) \right) d\rho \\ &= \exp \left(- \int_{S^{d-1}} |\theta \cdot s|^\alpha (1 - i \operatorname{sgn}(\theta \cdot s) \tan(\frac{\pi\alpha}{2})) \right) \Gamma(ds) \end{aligned}$$

where $S^{d-1} \subset \mathbb{R}^d$ is the unit sphere, $\theta \cdot s = \theta_1 s_1 + \dots + \theta_d s_d$, and $\Gamma(ds)$ is the finite measure on S^{d-1} obtained from $(\sum_i f_i(y)^2)^{\alpha/2} \rho(dy)$ by the transformation $y \mapsto (f_1(y), \dots, f_d(y)) / (\sum_i f_i(y)^2)^{1/2}$; see section 3.2 of [16]. Here Γ is the so-called spectral measure of the α -stable random vector $(I(f_1), \dots, I(f_d))$, which characterizes the joint distribution.

In our case E will be either \mathbb{R} or the interval $(0, 1]$. Then the above Poisson integrals can also be viewed as stochastic integrals using a Lévy process $L = L^\Xi$ with mean zero, constructed from the Poisson point process Ξ in the usual manner via compensation. Thus,

$$\int_E f dM = \int_E f dL. \quad (14)$$

2.4 An asymptotic expansion for the static case

Consider a beta coalescent back from time 0, and let $N_n(r)$ be the number of blocks in the partition at time r , so in particular $N_n(0) = n$. Let $R_0 = 0$, and for $k \geq 1$, let

$$R_k = R_{n,k} = \inf\{r > R_{k-1} : N_n(r) \neq N_n(R_{k-1})\}.$$

Let $\tau_n = \max\{k : R_k < \infty\}$. Thus, $R_1 < \dots < R_{\tau_n}$ are the times at which mergers occur, and τ_n is the number of mergers before only one block remains. For $0 \leq k \leq \tau_n$, let

$$X_k = X_{n,k} = N_n(R_k),$$

and let $X_k = 1$ for $k > \tau_n$, which means $n = X_0 > X_1 > \dots > X_{\tau_n} = 1$. The process $(X_k)_{k=0}^{\tau_n}$ is called the block-counting process associated with the beta coalescent.

The result below shows that certain functionals of the block-counting process have an asymptotic stable law as $n \rightarrow \infty$. For this purpose, we consider the beta coalescent, constructed from Υ_n as in Section 2.1, and the α -stable random measures $M_n = M^{\Theta_n}$ obtained from Poisson point processes Ξ_n as described in Section 2.3, with $E = (0, 1]$, intensity (6) and $\Xi_n = \Theta_n$ (arising from Υ_n as described in Section 2.2). In view of (11) the control measure of M_n is

$$\rho(dx) = \frac{\pi}{2 \sin(\frac{\pi\alpha}{2}) \Gamma(\alpha + 1)} \cdot \frac{\alpha(\alpha - 1)}{\Gamma(2 - \alpha)} dx = \frac{\pi(\alpha - 1)}{2 \sin(\frac{\pi\alpha}{2}) \Gamma(\alpha) \Gamma(2 - \alpha)}. \quad (15)$$

We denote by \mathcal{F} the set of all differentiable functions f which for some $c > 0$ and $0 < \zeta < \frac{1}{\alpha}$ obey

$$|f'(x)| \leq cx^{-\zeta-1} \quad (16)$$

for all $x \in (0, 1]$.

Theorem 2. Let $(X_k)_{k=0}^{\tau_n}$ be the block-counting process associated with the beta coalescent, and let $f \in \mathcal{F}$. Then as $n \rightarrow \infty$,

$$\frac{1}{\alpha-1} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) = n \int_0^1 f(x) dx - n^{1/\alpha} \int_{(0,1]} f(x) M_n(dx) + o_P(n^{1/\alpha}), \quad (17)$$

where M_n is an α -stable random measure with control measure (15).

The statement of Theorem 2 is an asymptotic version of the equation

$$\gamma \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) = n \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) \frac{Y_k}{n} - n^{1/\alpha} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) \frac{Y_k - \gamma}{n^{1/\alpha}}, \quad (18)$$

where $\gamma = 1/(\alpha-1)$ and $Y_k = X_k - X_{k+1}$, $0 \leq k < \tau_n$. The proof consists of showing that the sums appearing in the right-hand side of (18) may asymptotically be replaced by the integrals $\int_0^1 f(x) dx$ and $\int_{(0,1]} f(x) M_n(dx)$, respectively.

Since the control measure, and hence the distribution, of M_n does not depend on n , the asymptotic expansion (17) in Theorem 2 directly implies convergence in distribution not only for a single $f \in \mathcal{F}$, but also for finitely many f_1, \dots, f_d . To ease notation, let us define, for $f \in \mathcal{F}$ and $n \in \mathbb{N}$,

$$\mathcal{J}_n(f) := n^{-1/\alpha} \left(\frac{1}{\alpha-1} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) - n \int_0^1 f(x) dx \right). \quad (19)$$

Corollary 3. Let $f, f_1, \dots, f_d \in \mathcal{F}$. Then

$$\mathcal{J}_n(f) \Rightarrow S_\alpha(\sigma_f, -\beta_f, 0)$$

and

$$(\mathcal{J}_n(f_1), \dots, \mathcal{J}_n(f_d)) \Rightarrow (-I(f_1), \dots, -I(f_d)),$$

where σ_f, β_f are as in (13), $I(f)$ is as in (12) with an α -stable random measure M with control measure (15), and \Rightarrow denotes convergence in distribution as $n \rightarrow \infty$.

2.5 The evolving beta coalescent

Now we transform the asymptotic expansion from Theorem 2 to the evolving coalescent. Here it is convenient to use Lévy processes. Let

$$L_n = (L_{n,s})_{-\infty < s < \infty} := L^{\Psi_n}, \quad L'_n = (L'_{n,x})_{0 \leq x \leq 1} := L^{\Theta_n} \quad (20)$$

be the mean zero stable Lévy processes with $L_{n,0} = L'_{n,0} = 0$ a.s. and with jumps $\Delta L_{n,s} = u$ and $\Delta L'_{n,x} = y$ for all points (s, u) from Ψ_n and (x, y) from Θ_n , respectively. For the time-reversals we write

$$\begin{aligned} \hat{L}_{n,r} &:= L_{n,s-}, \quad s = -r, \quad r \geq 0, \\ \hat{L}'_{n,w} &:= L'_{n,x-} - L_{n,1}, \quad w = 1 - x, \quad 0 < x \leq 1. \end{aligned}$$

Then the mapping from (4) translates into the following lemma.

Lemma 4. Let $f : (0, 1] \rightarrow \mathbb{R}$ be such that $\int_0^1 |f(x)|^\alpha dx < \infty$. Then

$$\int_0^\infty f(m(r))m(r) d\hat{L}_{n,r} = \int_0^1 f(1-w) d\hat{L}'_{n,w}.$$

Proof. The assumption on f guarantees that the integrals are well-defined. The processes $(\hat{L}_{n,r})_{r \geq 0}$ and $(\hat{L}'_{n,m(r)})_{r \geq 0}$ are mean zero Lévy processes, and hence martingales with respect to the filtration

$$\mathcal{F}_r := \sigma(\Psi_n |_{[-r,0] \times \mathbb{R}^+}) = \sigma(\Theta_n |_{[m(r),1] \times \mathbb{R}^+}), \quad r \geq 0.$$

Consequently

$$J_r := \int_0^r f(m(q))m(q) d\hat{L}_{n,q} - \int_0^{1-m(r)} f(1-w) d\hat{L}'_{n,w}, \quad r \geq 0,$$

is a local martingale. Jumps can only occur at points (r, u) in Ψ_n ; because of (4) they vanish:

$$\Delta J_r = -f(m(r))m(r)u + f(x)y = 0.$$

Thus, J is a.s. continuous. Moreover, since the underlying processes are Lévy processes without a Brownian component, the quadratic variation of J is $[J]_\infty = \sum_{r \geq 0} (\Delta J_r)^2 = 0$ a.s. Thus $J_\infty = J_0 = 0$ a.s. \square

From (20) and (14) it follows that $M_n = M^{\Theta_n}$ satisfies

$$\int_{(0,1]} f(x) M_n(dx) = \int_0^1 f(x) dL'_{n,x} = - \int_0^1 f(1-w) d\hat{L}'_{n,w}.$$

Now we apply Lemma 4 to get

$$\int_{(0,1]} f(x) M_n(dx) = - \int_0^\infty f(m(r))m(r) dL_{n,-r}. \quad (21)$$

Let us now proceed to consider the evolving coalescent $(\mathcal{T}_n(t), t \in \mathbb{R})$ described in Section 2.1. For each $s \in \mathbb{R}$ and $n \in \mathbb{N}$, we denote the block counting process of the coalescent tree $\mathcal{T}_n(n^{1-\alpha}s)$ by $(X_k^s)_{k=0}^{\tau_n^s}$. By shifting the origin of the scaled time to the time point s and re-centering the process L at this new time origin (which does not affect its increments), we can apply Theorem 2 together with (21) and conclude that

$$\frac{1}{\alpha-1} \sum_{k < \tau_n^s} f\left(\frac{X_k^s}{n}\right) = n \int_0^1 f(x) dx + n^{1/\alpha} \int_0^\infty f(m(r))m(r) dL_{n,s-r} + o_P(n^{1/\alpha}).$$

Writing $\mathcal{J}_f^n(s)$ for the random variable (19) with $(X_k) = (X_k^0)$ replaced by (X_k^s) , we thus obtain

$$\mathcal{J}_f^n(s) = \int_0^\infty f(m(r))m(r) dL_{n,s-r} + o_P(1).$$

Since the distribution of the Poisson point process Ψ_n , and hence also that of the Lévy process $L_n = L^{\Psi_n}$, does not depend on n , we obtain the following result for the evolving beta coalescent.

Corollary 5. For $f \in \mathcal{F}$ and $s \in \mathbb{R}$, let $\mathcal{J}_{n,s}(f)$ be as in (19), but now evaluated at the coalescent tree $\mathcal{T}_n(n^{1-\alpha}s)$ instead of $\mathcal{T}_n(0)$. Then the sequence of stationary processes $(\mathcal{J}_{n,s}(f))_{-\infty < s < \infty}$, $n \geq 1$, converges as $n \rightarrow \infty$ in finite-dimensional distributions to the moving average process

$$\int_0^\infty f(m(r))m(r) dL_{s-r}, \quad -\infty < s < \infty,$$

where m is given by (5) and $(L_s)_{-\infty < s < \infty}$ is a mean zero Lévy process with $L_0 = 0$ and Lévy measure given by (2).

To understand better how these functionals of the beta coalescent evolve over time, note that the stable random variable $\int_{(0,1]} f(x) M_n(dx)$ from Theorem 2, which gives the limit of the functional $\mathcal{J}_{n,0}(f)$, is a function of the Poisson process Θ_n , and is therefore also a function of the Poisson process Ψ_n . Likewise, the stable random variable that gives the limit of $\mathcal{J}_{n,s}(f)$ can be expressed as a function of a Poisson process Θ_n^s and as a function of a Poisson process Ψ_n^s . The Poisson process Ψ_n^s can be obtained from Ψ_n by a simple time shift. If $(-r, u)$ is a point of Ψ_n , then $(-r-s, u)$ is a point of Ψ_n^s . We obtain the Poisson process Θ_n^s by applying the transformation $(s-r, u) \mapsto (x, y) = (m(r), m(r)u)$ to the points of $\Psi_n|_{(-\infty, s]}$, where $m(r)$ is again defined by (5). See Figure 3 for an illustration of how these point process evolve as s increases.

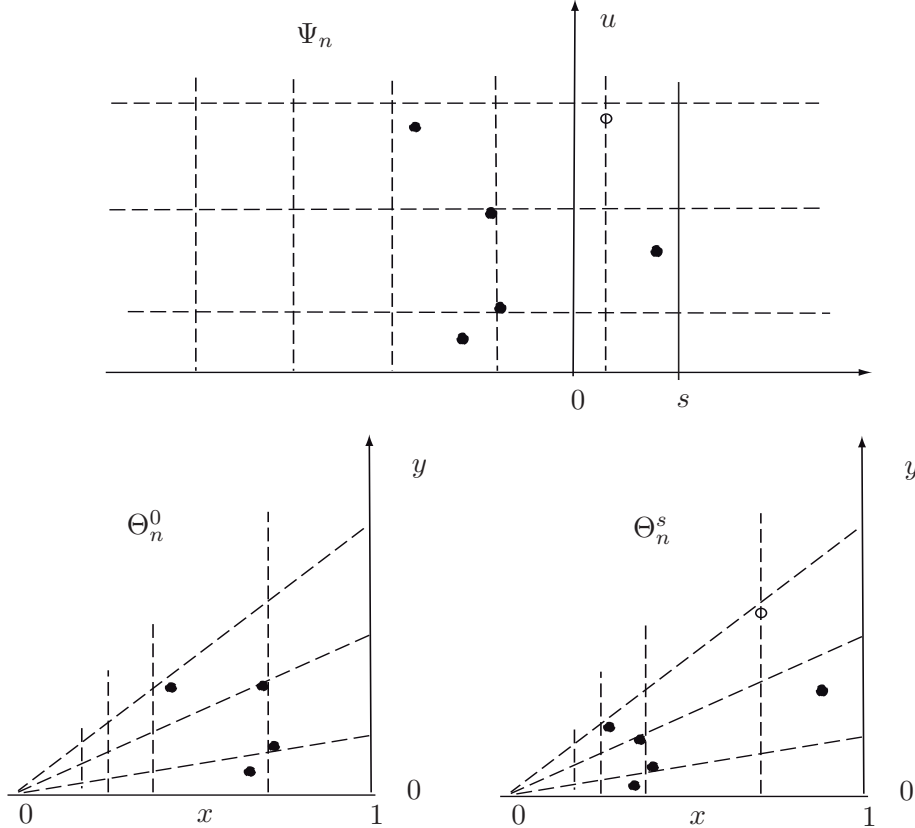


Figure 3: The point process Θ_n^s arises from $\Psi_n|_{(-\infty, s] \times \mathbb{R}^+}$ through the transformation $(s-r, u) \mapsto (x, y) = (m(r), m(r)u)$. As s increases above 0, the points of Θ_n^0 wander down to the left towards $(0, 0)$, and new points wander in from the right.

2.6 Functionals of the beta coalescent

In this section, we consider three functionals of the beta coalescent: the number of collisions, the total branch length, and the total length of external branches. We observe how Theorem 2 allows us to recover known results for the asymptotic distributions of these quantities for the static beta coalescent. Then Corollary 5 allows us to describe how these functionals behave over time in the evolving beta coalescent. We also obtain a new result about the ratio of the external branch length to the total branch length, which could be of interest for biological applications.

Example 6. Consider the number τ_n of collisions before just a single block remains. Because

$$\tau_n = \sum_{k < \tau_n} 1 = \frac{1}{\alpha - 1} \sum_{k < \tau_n} (\alpha - 1),$$

we can apply directly the result of Theorem 2 with $f(x) = \alpha - 1$ for all $x \in (0, 1]$. We get that for $1 < \alpha < 2$,

$$n^{-1/\alpha}(\tau_n - (\alpha - 1)n) \Rightarrow S(\sigma_1, -1, 0), \quad (22)$$

where

$$\sigma_1 = \left(\frac{\pi(\alpha - 1)^{1+\alpha}}{2 \sin(\frac{\pi\alpha}{2}) \Gamma(\alpha) \Gamma(2 - \alpha)} \right)^{1/\alpha}.$$

This agrees with the result of Lemma 4 in [5], where the limit on the right-hand side of (22) is expressed as $c_1 Z$ for $-Z$ satisfying (9) and $c_1 = (\alpha - 1)^{1+1/\alpha} / \Gamma(2 - \alpha)^{1/\alpha}$. This result had also been shown in [6, 7, 11], and the equivalence between the two ways of expressing the limit can be seen from (10).

Because we use this result in our proof of Theorem 2, we have not obtained here another independent proof of this result. The benefit is that our approach allows us to examine the common distribution of τ_n and other functionals. Also, Corollary 5 with $f(x) = \alpha - 1$ allows us to understand how the total number of collisions changes over time for the evolving beta coalescent. In particular, we see that the limit process is a stationary stable process that can be expressed, in a relatively simple way, as a moving average process.

Example 7. Consider next the total length \mathcal{L}_n of all branches in the coalescent tree. This quantity is of interest in Biology because the total branch length should be approximately proportional to the number of mutations observed in a sample of n individuals. Note that

$$\mathcal{L}_n = \sum_{k < \tau_n} X_k (R_{k+1} - R_k).$$

Define also

$$\mathcal{L}'_n = \sum_{k < \tau_n} \frac{X_k}{\lambda_{X_k}}, \quad \mathcal{L}''_n = \alpha \Gamma(\alpha) \sum_{k < \tau_n} X_k^{1-\alpha}.$$

Lemma 2.2 in [6] implies that as $m \rightarrow \infty$,

$$\lambda_m = \frac{1}{\alpha \Gamma(\alpha)} m^\alpha + O(m^{\alpha-1}). \quad (23)$$

Therefore, there is a constant $c > 0$ such that

$$|\mathcal{L}'_n - \mathcal{L}''_n| \leq \sum_{k < \tau_n} X_k \left| \frac{1}{\lambda_{X_k}} - \frac{\alpha \Gamma(\alpha)}{X_k^\alpha} \right| \leq \sum_{m=1}^n m \left| \frac{1}{\lambda_m} - \frac{\alpha \Gamma(\alpha)}{m^\alpha} \right| \leq \sum_{m=1}^n m \cdot c m^{-1-\alpha} = O(1). \quad (24)$$

Also, conditional on $\sigma(X) = \sigma(X_0, X_1, \dots, X_{\tau_n})$, the distribution of $X_k(R_{k+1} - R_k)$ is exponential with rate parameter λ_{X_k} . It follows that

$$E(\mathcal{L}_n - \mathcal{L}'_n \mid \sigma(X)) = 0 \quad (25)$$

and

$$\text{Var}(\mathcal{L}_n - \mathcal{L}'_n \mid \sigma(X)) = \sum_{k=0}^{n-1} X_k^2 1_{k < \tau_n} \cdot \frac{1}{\lambda_{X_k}^2} \leq \sum_{m=1}^n \frac{m^2}{\lambda_m^2} = O(1 \vee n^{3-2\alpha}). \quad (26)$$

It now follows from (24), (25), (26), and Chebyshev's Inequality that if $1 < \alpha < \frac{1}{2}(1 + \sqrt{5})$, so that $1 + \alpha - \alpha^2 > 0$, we have

$$n^{\alpha-1-1/\alpha}(\mathcal{L}_n - \mathcal{L}''_n) \Rightarrow 0.$$

Therefore, we may replace \mathcal{L}_n by \mathcal{L}''_n in asymptotic calculations. Because

$$n^{\alpha-1}\mathcal{L}''_n = \frac{1}{\alpha-1} \sum_{k < \tau_n} \alpha(\alpha-1)\Gamma(\alpha) \left(\frac{X_k}{n} \right)^{1-\alpha},$$

when $1 < \alpha < \frac{1}{2}(1 + \sqrt{5})$ we can apply Theorem 2 with $f(x) = \alpha(\alpha-1)\Gamma(\alpha)x^{1-\alpha}$. We get

$$n^{\alpha-1-1/\alpha} \left(\mathcal{L}_n - \frac{\alpha(\alpha-1)\Gamma(\alpha)n^{2-\alpha}}{2-\alpha} \right) \Rightarrow S_\alpha(\sigma_2, -1, 0),$$

where

$$\sigma_2 = \left(\frac{\pi \alpha^\alpha (\alpha-1)^{1+\alpha} \Gamma(\alpha)^{\alpha-1}}{2 \sin(\frac{\pi \alpha}{2}) \Gamma(2-\alpha) (1+\alpha-\alpha^2)} \right)^{1/\alpha},$$

which agrees with part (i) of Theorem 1 in [9]. It also follows from Corollary 5 that for the evolving beta coalescent, the evolution of the total branch length, scaled as above, converges in the sense of finite-dimensional distributions to a stationary stable process.

Example 8. Consider also the total length ℓ_n of all external branches in the tree. This quantity is also of interest in Biology, as it should be approximately proportional to the number of mutations that appear on just one individual in a sample of n individuals. It is shown in the proof of Theorem 1 in [5] that

$$\ell_n = \alpha(\alpha-1)^2 \Gamma(\alpha) n^{2-\alpha} + \alpha(2-\alpha) \Gamma(\alpha) n^{1-\alpha} \tau_n + o_P(n^{1+1/\alpha-\alpha}).$$

Therefore,

$$n^{\alpha-1} \ell_n - \alpha(\alpha-1)^2 \Gamma(\alpha) n = \frac{1}{\alpha-1} \sum_{k < \tau_n} \alpha(\alpha-1)(2-\alpha) \Gamma(\alpha) + o_P(n^{1/\alpha}),$$

so for $1 < \alpha < 2$, we can apply Theorem 2 with $f(x) = \alpha(\alpha-1)(2-\alpha)\Gamma(\alpha)$ to get

$$n^{\alpha-1-1/\alpha}(\ell_n - \alpha(\alpha-1)\Gamma(\alpha)n^{2-\alpha}) \Rightarrow S_\alpha(\sigma_3, -1, 0),$$

where

$$\sigma_3 = \left(\frac{\pi \alpha^\alpha (\alpha - 1)^{1+\alpha} (2 - \alpha)^\alpha \Gamma(\alpha)^{\alpha-1}}{2 \sin(\frac{\pi \alpha}{2}) \Gamma(2 - \alpha)} \right)^{1/\alpha},$$

in agreement with Theorem 1 of [5].

Example 9. Finally, we consider the quantity ℓ_n/\mathcal{L}_n , which in the biological setting should be approximately equal to the proportion of mutations that appear on only one individual. This ratio is potentially useful for drawing inferences about the genealogy of a population from data, in part because the value that we expect for this ratio does not depend on the mutation rate, which is often unknown. Indeed, it follows from results in [1] that the parameter α in the beta coalescent can be consistently estimated by the quantity $2 - \ell_n/\mathcal{L}_n$.

Assume that $1 < \alpha < \frac{1}{2}(1 + \sqrt{5})$. From the discussion in Examples 7 and 8, we see that

$$n^{\alpha-1-1/\alpha} \mathcal{L}_n = \frac{\alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha} n^{1-1/\alpha} - Z_1 + o_P(1),$$

where

$$Z_1 = \int_0^1 \alpha(\alpha-1)\Gamma(\alpha) x^{1-\alpha} M_n(dx),$$

and likewise

$$n^{\alpha-1-1/\alpha} \ell_n = \alpha(\alpha-1)\Gamma(\alpha) n^{1-1/\alpha} - Z_2 + o_P(1),$$

where

$$Z_2 = \int_0^1 \alpha(\alpha-1)(2-\alpha)\Gamma(\alpha) M_n(dx).$$

Therefore,

$$\begin{aligned} \frac{\ell_n}{\mathcal{L}_n} &= \frac{\alpha(\alpha-1)\Gamma(\alpha) - n^{-1+1/\alpha} Z_2 + o_P(n^{-1+1/\alpha})}{\frac{\alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha} - n^{-1+1/\alpha} Z_1 + o_P(n^{-1+1/\alpha})} \\ &= (2-\alpha) + n^{-1+1/\alpha} \left(\frac{(2-\alpha)^2}{\alpha(\alpha-1)\Gamma(\alpha)} Z_1 - \frac{2-\alpha}{\alpha(\alpha-1)\Gamma(\alpha)} Z_2 \right) + o_P(n^{-1+1/\alpha}) \\ &= (2-\alpha) + n^{-1+1/\alpha} \int_0^1 (2-\alpha)^2 (x^{1-\alpha} - 1) M_n(dx) + o_P(n^{-1+1/\alpha}). \end{aligned}$$

It follows that

$$n^{1-1/\alpha} \left(\frac{\ell_n}{\mathcal{L}_n} - (2-\alpha) \right) \Rightarrow S_\alpha(\sigma_4, 1, 0),$$

where

$$\sigma_4 = (2-\alpha)^2 \left(\frac{\pi(\alpha-1)}{2 \sin(\frac{\pi \alpha}{2}) \Gamma(\alpha) \Gamma(2-\alpha)} \int_0^1 (x^{1-\alpha} - 1)^\alpha dx \right)^{1/\alpha}.$$

Using the substitution $y = x^{\alpha-1}$, the integral transforms to a beta integral:

$$\int_0^1 (x^{1-\alpha} - 1)^\alpha dx = \frac{1}{\alpha-1} \int_0^1 (1-y)^\alpha y^{\frac{2-\alpha^2}{\alpha-1}} dy = \frac{\Gamma(\alpha+1) \Gamma(\frac{\alpha+1-\alpha^2}{\alpha-1})}{(\alpha-1) \Gamma(\frac{\alpha}{\alpha-1})}.$$

Altogether,

$$\sigma_4 = (2-\alpha)^2 \left(\frac{\pi \alpha}{2 \sin(\frac{\pi \alpha}{2}) \Gamma(2-\alpha) \Gamma(\frac{\alpha}{\alpha-1})} \right)^{1/\alpha}.$$

If $\frac{1}{2}(1 + \sqrt{5}) \leq \alpha < 2$, then the fluctuations in \mathcal{L}_n are of a higher order of magnitude than the fluctuations of ℓ_n , so the asymptotic distribution of ℓ_n/\mathcal{L}_n is determined by the asymptotics of \mathcal{L}_n given in Theorem 2 of [5]. In particular, when $\frac{1}{2}(1 + \sqrt{5}) < \alpha < 2$, the asymptotic distribution of ℓ_n/\mathcal{L}_n is no longer a stable law.

3 Proofs

Let us remark in advance that for f satisfying (16) with $\zeta < 2$ by linearity we may and will assume the following properties: $f(x) \geq 1$ for all $x \in (0, 1]$, $f(x)$ is monotonically decreasing and $x^2 f(x)$ is monotonically increasing.

Indeed, for any f satisfying the condition $|f'(x)| \leq cx^{-\zeta-1}$, $x \in (0, 1]$, we may write $f = f_1 - f_2$ with

$$f_1(x) = f(x) + 3c\zeta^{-1}x^{-\zeta} + d, \quad f_2(x) = 3c\zeta^{-1}x^{-\zeta} + d.$$

Then f_1 and f_2 fulfil these three requirements, if we let $d = 1 - f(1) \wedge 0$. For f_2 this is obvious, since $\zeta < 2$. Furthermore $f'_1(x) \leq -2cx^{-\zeta-1}$. This implies that f_1 is decreasing and also $f_1(x) \geq 1$ for all x , since $f_1(1) \geq 1$. Moreover $d + f(1) \geq 0$, thus $f_1(1) \geq 3c\zeta^{-1}$ and

$$f_1(x) = f_1(1) - \int_x^1 f'_1(y) dy \geq 2c\zeta^{-1} + \int_x^1 2cy^{-\zeta-1} dy = 2c\zeta^{-1}x^{-\zeta} \geq 2cx^{-\zeta}.$$

Taking also into account $f'_1(x) \geq -4cx^{-\zeta-1}$ we obtain

$$\frac{d}{dx}x^2 f_1(x) = 2x f_1(x) + x^2 f'_1(x) \geq 0.$$

This gives the assertion.

Note also for f satisfying (16), there is a positive constant c such that

$$f(x) \leq cx^{-\zeta} \tag{27}$$

for all $x \in (0, 1]$ and therefore

$$\int_0^1 f(x)^\alpha dx < \infty.$$

To facilitate notation we adopt, here and throughout the rest of the paper, the convention that $c > 0$ denotes a constant, only dependent on α , which may change its value from term to term.

3.1 The number of blocks for the beta coalescent

We assemble here some results about the evolution of the number of blocks for the beta coalescent. We adopt the notation of Section 2.4, so that τ_n is the total number of mergers and $(X_k)_{k=0}^{\tau_n}$ is the block counting process. Let

$$Y_k = X_k - X_{k+1}, \quad k \geq 0,$$

which are the numbers of blocks lost during the mergers.

Define

$$q_i = \frac{\alpha}{\Gamma(2 - \alpha)} \frac{\Gamma(i + 1 - \alpha)}{\Gamma(i + 2)}, \quad i \geq 1.$$

The numbers q_i are the weights of a probability distribution on \mathbb{N} (see [9]). From Stirling's formula

$$q_i = \frac{\alpha}{\Gamma(2-\alpha)} i^{-\alpha-1} (1 + o(1)) \quad (28)$$

for $i \rightarrow \infty$. Let

$$\gamma = \sum_{i \geq 1} i q_i = \frac{1}{\alpha - 1}, \quad (29)$$

where the last equality is formula (5) in [9]. See also [3, 6] for a discussion of this probability distribution.

It has been known since the work of Bertoin and Le Gall [3] that the distribution of the random variables Y_0, Y_1, \dots is well approximated by $(q_i)_{i=1}^\infty$. The next result gives a bound on the accuracy of this approximation.

Lemma 10. *There is a number $c < \infty$ such that for $j \geq 2$ and $1 \leq i < j$*

$$|P(Y_0 = i \mid X_0 = j) - q_i| \leq \frac{c i q_i}{j}$$

and thus

$$P(Y_0 = i \mid X_0 = j) \leq (1 + c) q_i.$$

Proof. In the proof of Lemma 3 in [9] (see there the two displayed formulas before (9)) it is shown that there are real numbers b_j such that for $1 \leq i < j$

$$\left(1 - \frac{i}{j}\right) q_i \leq b_j P(Y_0 = i \mid X_0 = j) \leq q_i, \quad 1 - \frac{\gamma}{j} \leq b_j \leq 1,$$

and so for $j > \gamma$

$$\left(1 - \frac{i}{j}\right) q_i \leq P(Y_0 = i \mid X_0 = j) \leq \frac{1}{1 - \frac{\gamma}{j}} q_i.$$

This gives our claim in the case when $j > 2\gamma$. The other finitely many cases are covered too, if we choose c sufficiently large. \square

The next two lemmas contain our first applications of these estimates.

Lemma 11. *For $n \rightarrow \infty$*

$$\max_{k < \tau_n} \frac{X_k}{X_{k+1}} = O_P(1).$$

Proof. Let $a > 1$ and $\eta = (a - 1)/a$. Because $Y_k = X_k - X_{k+1}$,

$$P\left(\max_{k < \tau_n} \frac{X_k}{X_{k+1}} > a\right) = P\left(\max_{k < \tau_n} \frac{Y_k}{X_k} > \eta\right) \leq E\left(\sum_{k < \tau_n} 1_{\{Y_k > \eta X_k\}}\right).$$

From Lemma 10 and (28)

$$P(Y_0 > \eta j \mid X_0 = j) \leq c \sum_{\eta j < i < j} i^{-\alpha-1} \leq c((\eta j)^{-\alpha} - j^{-\alpha})$$

and consequently, using that $\tau_n \leq n - 1$,

$$P\left(\max_{k < \tau_n} \frac{X_k}{X_{k+1}} > a\right) \leq \sum_{k=0}^{n-2} P(Y_k > \eta X_k) \leq c(\eta^{-\alpha} - 1) \sum_{k=0}^{n-2} E(X_k^{-\alpha}) \leq c(\eta^{-\alpha} - 1) \sum_{j=1}^{\infty} \frac{1}{j^\alpha}.$$

Since $\alpha > 1$, the series is convergent. Also as $a \rightarrow \infty$, we have $\eta \rightarrow 1$, and the claim follows. \square

Next let us introduce the stopping times

$$\tau_n(a) := \min\{k \geq 0 : X_k \leq an\}, \quad 0 < a \leq 1.$$

In particular, $\tau_n(1) = 0$.

Lemma 12. *For $n \rightarrow \infty$*

$$\tau_n(a) = \frac{(1-a)n}{\gamma} + o_P(n).$$

Proof. Let $\xi \in (\alpha^{-1}, 1)$. Then for $k < \tau_n$ from Lemma 10 and (28)

$$P(Y_k > n^\xi \mid X_k) \leq c \sum_{i > n^\xi} i^{-\alpha-1} \leq c(n^\xi)^{-\alpha}$$

and since $\tau_n \leq n - 1$

$$P\left(\max_{k < \tau_n} Y_k > n^\xi\right) \leq \sum_{k < n-1} P(Y_k > n^\xi) \leq cn^{1-\xi\alpha} = o(1)$$

or $\max_{k < \tau_n} Y_k = O_P(n^\xi)$. Thus $X_{\tau_n(a)-1} - X_{\tau_n(a)} = Y_{\tau_n(a)-1} = O_P(n^\xi)$. Also by definition $X_{\tau_n(a)} \leq an < X_{\tau_n(a)-1}$. Since $\xi < 1$ this gives

$$n^{-1}X_{\tau_n(a)} \rightarrow a$$

in probability.

Observe that $\tau_n - \tau_n(a) = \tau_{X_{\tau_n(a)}}$, and recall that it was shown in [6, 7, 11] that

$$\tau_n = n/\gamma + o_P(n).$$

Therefore, using the strong Markov property, we obtain

$$\tau_n - \tau_n(a) = \frac{X_{\tau_n(a)}}{\gamma} + o_P(X_{\tau_n(a)}) = \frac{an}{\gamma} + o_P(n),$$

which gives the claim. \square

The next lemma pertains to the evolution of the number of blocks in continuous time and follows fairly directly from results in [2], where the number of blocks was studied for the beta coalescent started with infinitely many blocks. Recall the definition of $m(r)$ from (5) and let as above $N_n(r)$ denote the number of blocks at time r .

Lemma 13. *Consider the beta coalescent started with n blocks at time zero. Let $\varepsilon > 0$. Suppose $h : \mathbb{N} \rightarrow (0, \infty)$ is a function such that $\lim_{n \rightarrow \infty} n^{1-\alpha}h(n) = 0$. Then*

$$\lim_{n \rightarrow \infty} P((1-\varepsilon)m(r)n \leq N_n(n^{1-\alpha}r) \leq (1+\varepsilon)m(r)n \text{ for all } r \in [0, h(n)]) = 1.$$

Proof. Consider a beta coalescent started with infinitely many blocks at time zero, and let $N(r) = N_\infty(r)$ denote the number of blocks at time r . Theorem 1.1 of [2] states that

$$\lim_{r \downarrow 0} r^{1/(\alpha-1)} N(r) = (\alpha\Gamma(\alpha))^{1/(\alpha-1)} \text{ a.s.} \quad (30)$$

The strategy of the proof will be to bound the process $(N_n(r), r \geq 0)$ from above by the process $(N(r), r \geq 0)$ started at time $(1-\delta)\alpha\Gamma(\alpha)n^{1-\alpha}$, when there will typically be more than n blocks, and from below by the process $(N(r), r \geq 0)$ started at time $(1+\delta)\alpha\Gamma(\alpha)n^{1-\alpha}$, when there will typically be fewer than n blocks.

Choose $\delta > 0$ sufficiently small that

$$(1+\delta) \left(\frac{\alpha\Gamma(\alpha)}{(1-\delta)\alpha\Gamma(\alpha) + r} \right)^{1/(\alpha-1)} \leq (1+\varepsilon)m(r) \quad (31)$$

and

$$(1-\delta) \left(\frac{\alpha\Gamma(\alpha)}{(1+\delta)\alpha\Gamma(\alpha) + r} \right)^{1/(\alpha-1)} \geq (1-\varepsilon)m(r). \quad (32)$$

For $r > 0$, define the event

$$G(r) = \{(1-\delta)(\alpha\Gamma(\alpha))^{1/(\alpha-1)} r^{-1/(\alpha-1)} n \leq N(n^{1-\alpha}r) \leq (1+\delta)(\alpha\Gamma(\alpha))^{1/(\alpha-1)} r^{-1/(\alpha-1)} n\}.$$

For $n \in \mathbb{N}$, let $g(n) = (1+\delta)(\alpha\Gamma(\alpha) + h(n))$. Then $\lim_{n \rightarrow \infty} g(n)n^{1-\alpha} = 0$, so (30) gives

$$\lim_{n \rightarrow \infty} P(G(r) \text{ occurs for all } r \in [0, g(n)]) = 1 \quad (33)$$

and

$$\lim_{n \rightarrow \infty} P(N((1+\delta)\alpha\Gamma(\alpha)n^{1-\alpha}) \leq n \leq N((1-\delta)\alpha\Gamma(\alpha)n^{1-\alpha})) = 1. \quad (34)$$

Therefore,

$$\begin{aligned} & P(N_n(n^{1-\alpha}r) > (1+\varepsilon)m(r)n \text{ for some } r \in [0, h(n)]) \\ & \leq P(N((1-\delta)\alpha\Gamma(\alpha)n^{1-\alpha}) < n) \\ & \quad + P(N(((1-\delta)\alpha\Gamma(\alpha) + r)n^{1-\alpha}) > (1+\varepsilon)m(r)n \text{ for some } r \in [0, h(n)]), \end{aligned}$$

which tends to zero as $n \rightarrow \infty$ by (31), (33), and (34). Likewise,

$$\begin{aligned} & P(N_n(n^{1-\alpha}r) < (1-\varepsilon)m(r)n \text{ for some } r \in [0, h(n)]) \\ & \leq P(N((1+\delta)\alpha\Gamma(\alpha)n^{1-\alpha}) > n) \\ & \quad + P(N(((1+\delta)\alpha\Gamma(\alpha) + r)n^{1-\alpha}) < (1-\varepsilon)m(r)n \text{ for some } r \in [0, h(n)]), \end{aligned}$$

which tends to zero as $n \rightarrow \infty$ by (32), (33), and (34). The result follows. \square

3.2 Functionals of the block counting process

We again use the notation of section 2.4, so that $n = X_0 > X_1 > \dots > X_{\tau_n} = 1$ is the block-counting process of the beta n -coalescent and $Y_k = X_k - X_{k+1}$.

Proposition 14. Let $f : (0, 1] \rightarrow \mathbb{R}$ be a positive, decreasing, differentiable function such that

$$|f'(x)| \leq cx^{-\zeta-1}$$

for some $c > 0$. If $\zeta < 1$, then f is integrable and

$$\gamma \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) = n \int_0^1 f(x) dx + o_P(n).$$

If moreover $\zeta < 1/\alpha$, i.e. $f \in \mathcal{F}$, then

$$\gamma \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) = n \int_0^1 f(x) dx - \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right)(Y_k - \gamma) + o_P(n^{1/\alpha}).$$

Proof. Choose $\varepsilon > 0$. Let $0 = a_0 < a_1 < \dots < a_m = 1$ be a partition of $[0, 1]$. Then by the assumptions on f

$$0 \leq \sum_{\tau_n(a_1) \leq k < \tau_n} f\left(\frac{X_k}{n}\right) \leq \sum_{1 \leq j \leq a_1 n} f\left(\frac{j}{n}\right) \leq n \int_0^{a_1} f(x) dx \leq \frac{\varepsilon}{3}n,$$

if only a_1 is sufficiently small.

Also for $i > 1$ by monotonicity and Lemma 12

$$\sum_{\tau_n(a_i) \leq k < \tau_n(a_{i-1})} f\left(\frac{X_k}{n}\right) \geq f(a_i)(\tau_n(a_{i-1}) - \tau_n(a_i)) = f(a_i) \frac{(a_i - a_{i-1})n}{\gamma} + o_P(n)$$

which implies that

$$\sum_{k < \tau_n(a_1)} f\left(\frac{X_k}{n}\right) \geq \frac{n}{\gamma} \int_{a_1}^1 f(x) dx - \frac{\varepsilon}{3}n + o_P(n),$$

if only the partition is chosen fine enough. Combining the estimates we obtain

$$P\left(\sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) < \frac{n}{\gamma} \int_0^1 f(x) dx - \varepsilon n\right) = o(1).$$

In the same manner we may bound $\sum_{k < \tau_n} f(X_k/n)$ from above such that the first claim follows.

As to the second one, from a Taylor expansion with $X_{k+1} \leq \bar{X}_k \leq X_k$,

$$\int_{X_{k+1}/n}^{X_k/n} f(x) dx = f\left(\frac{X_k}{n}\right) \frac{Y_k}{n} + \frac{1}{2} f'\left(\frac{\bar{X}_k}{n}\right) \left(\frac{Y_k}{n}\right)^2.$$

Therefore R_n , given by

$$\gamma \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) = n \int_{1/n}^1 f(x) dx - \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right)(Y_k - \gamma) + R_n,$$

fulfils by assumption

$$|R_n| \leq \sum_{k=0}^{\tau_n-1} \left| f'\left(\frac{\bar{X}_k}{n}\right) \right| \frac{Y_k^2}{n} \leq c \sum_{k=0}^{\tau_n-1} \left(\frac{X_{k+1}}{n}\right)^{-\zeta-1} \frac{Y_k^2}{n}$$

or

$$|R_n| \leq cn^\zeta \left(\max_{k < \tau_n} \frac{X_k}{X_{k+1}} \right)^{\zeta+1} \sum_{k=0}^{\tau_n-1} X_k^{-\zeta-1} Y_k^2.$$

Because of Lemma 10, equation (28), and the fact that $\alpha < 2$, for $k < \tau_n$ we get

$$E(Y_k^2 \mid X_k) \leq c \sum_{i < X_k} i^2 q_i \leq c X_k^{2-\alpha}$$

and therefore

$$E\left(\sum_{k < \tau_n} X_k^{-\zeta-1} Y_k^2\right) \leq c E\left(\sum_{k < \tau_n} X_k^{1-\alpha-\zeta}\right) \leq c \sum_{j \geq 1} j^{1-\alpha-\zeta}.$$

Furthermore $1 - \alpha - \zeta = \frac{1}{\alpha} - \zeta - \frac{1}{\alpha}(\alpha - 1)^2 - 1 < -1$, if only ζ is chosen sufficiently close to $1/\alpha$, such that the right-hand series is convergent. Altogether in view of Lemma 11 we obtain

$$R_n = O_P(n^\zeta) = o_P(n^{1/\alpha}).$$

Finally, since $\zeta < 1/\alpha < 1$ it follows from (27) that

$$n \int_0^{1/n} f(x) dx \leq cn \left(\frac{1}{n}\right)^{1-\zeta} = cn^\zeta = o(n^{1/\alpha}).$$

This gives the second assertion. □

Lemma 15. *Let $f \in \mathcal{F}$. Then for any $\eta > 0$ there is an $\varepsilon > 0$ such that for all $n \geq 1$*

$$P\left(\left|d + n^{-1/\alpha} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right)(Y_k 1_{\{f(X_k/n)Y_k \leq \varepsilon n^{1/\alpha}\}} - \gamma)\right| > \eta\right) \leq \eta$$

with

$$d = d(\varepsilon) = \varepsilon^{1-\alpha} \frac{\alpha}{\Gamma(2-\alpha)} \int_0^1 f(x)^\alpha dx. \quad (35)$$

Proof. Suppressing the dependence on n in the notation, let

$$A_k = \{f(X_k/n)Y_k \leq \varepsilon n^{1/\alpha}\},$$

$$\gamma(j) = E(Y_0 1_{A_0} \mid X_0 = j).$$

Since (X_k) is a Markov chain and τ_n is a stopping time, the random variables

$$f\left(\frac{X_k}{n}\right)(Y_k 1_{A_k} - \gamma(X_k)) 1_{\{k < \tau_n\}}$$

have zero mean and are uncorrelated. Therefore,

$$\text{Var}\left(\sum_{k < \tau_n} f\left(\frac{X_k}{n}\right)(Y_k 1_{A_k} - \gamma(X_k))\right) \leq \sum_{k=0}^{n-1} E\left(f\left(\frac{X_k}{n}\right)^2 (Y_k 1_{A_k} - \gamma(X_k))^2 1_{\{k < \tau_n\}}\right).$$

From Lemma 10 and (28), we see that

$$E((Y_0 1_{A_0} - \gamma(X_0))^2 \mid X_0 = j) \leq E(Y_0^2 1_{A_0} \mid X_0 = j) \leq c \sum_{i \leq \varepsilon n^{1/\alpha}/f(j/n)} i^2 q_i \leq c \left(\frac{\varepsilon n^{1/\alpha}}{f(j/n)} \right)^{2-\alpha}.$$

Thus,

$$\text{Var} \left(\sum_{k < \tau_n} f \left(\frac{X_k}{n} \right) (Y_k 1_{A_k} - \gamma(X_k)) \right) \leq c \varepsilon^{2-\alpha} n^{2/\alpha-1} \sum_{k=0}^{n-1} E \left(f \left(\frac{X_k}{n} \right)^\alpha \right).$$

Because $\zeta < 1/\alpha$, we have, using (27),

$$\sum_{k=0}^{n-1} E \left(f \left(\frac{X_k}{n} \right)^\alpha \right) \leq \sum_{j=1}^n f \left(\frac{j}{n} \right)^\alpha \leq c \sum_{j=1}^n \left(\frac{j}{n} \right)^{-\alpha\zeta} \leq cn,$$

and it follows that

$$\text{Var} \left(n^{-1/\alpha} \sum_{k < \tau_n} f \left(\frac{X_k}{n} \right) (Y_k 1_{A_k} - \gamma(X_k)) \right) \leq c \varepsilon^{2-\alpha}.$$

Thus, by Chebyshev's Inequality, if ε is sufficiently small, then

$$P \left(\left| n^{-1/\alpha} \sum_{k < \tau_n} f \left(\frac{X_k}{n} \right) (Y_k 1_{A_k} - \gamma(X_k)) \right| > \eta \right) \leq \eta. \quad (36)$$

It remains to replace $\gamma(X_k)$ by γ in this formula. From

$$\gamma(j) = \sum_{i \leq \varepsilon n^{1/\alpha}/f(j/n)} i P(Y_0 = i \mid X_0 = j)$$

we get from Lemma 10 and (28) the estimate, uniform in n and k ,

$$\left| \gamma(j) - \sum_{i \leq (j-1) \wedge \varepsilon n^{1/\alpha}/f(j/n)} i q_i \right| \leq \frac{c}{j} \sum_{i \leq j-1} i^2 q_i = O(j^{1-\alpha}).$$

From (28) and (29),

$$\sum_{i \leq (j-1) \wedge \varepsilon n^{1/\alpha}/f(j/n)} i q_i = \gamma - \frac{\alpha}{(\alpha-1)\Gamma(2-\alpha)} \left(\frac{\varepsilon n^{1/\alpha}}{f(j/n)} \right)^{1-\alpha} (1 + o(1)) + O(j^{1-\alpha}),$$

where the $o(1)$ goes to 0 with n going to infinity, uniformly in j . Putting these formulas together we arrive at

$$\gamma(j) = \gamma - \frac{\alpha}{(\alpha-1)\Gamma(2-\alpha)} \varepsilon^{1-\alpha} n^{-1+1/\alpha} f(j/n)^{\alpha-1} (1 + o(1)) + O(j^{1-\alpha}).$$

It follows that

$$\begin{aligned} \sum_{k < \tau_n} f \left(\frac{X_k}{n} \right) \gamma(X_k) &= \gamma \sum_{k < \tau_n} f \left(\frac{X_k}{n} \right) - \frac{\alpha + o(1)}{(\alpha-1)\Gamma(2-\alpha)} \varepsilon^{1-\alpha} n^{-1+1/\alpha} \sum_{k < \tau_n} f \left(\frac{X_k}{n} \right)^\alpha \\ &\quad + O \left(\sum_{k < \tau_n} f \left(\frac{X_k}{n} \right) X_k^{1-\alpha} \right). \end{aligned} \quad (37)$$

Now, since $1 - \alpha - \zeta = \frac{1}{\alpha} - \zeta - \frac{1}{\alpha}(\alpha - 1)^2 - 1 < -1$, if ζ is sufficiently close to $1/\alpha$,

$$\sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) X_k^{1-\alpha} \leq cn^\zeta \sum_{k < \tau_n} X_k^{1-\alpha-\zeta} \leq cn^\zeta \sum_{j \geq 1} j^{1-\alpha-\zeta} = O(n^\zeta) = o(n^{1/\alpha}).$$

Because $|\frac{d}{ds} f(s)^\alpha| = \alpha f(s)^{\alpha-1} |f'(s)| \leq cs^{-\zeta(\alpha-1)} s^{-\zeta-1} = cs^{-\alpha\zeta-1}$ and $\alpha\zeta < 1$, we may estimate the middle term in the right-hand side of (37) by applying the first statement of Proposition 14. This implies

$$\sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) \gamma(X_k) = \gamma \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) - \frac{\alpha}{\Gamma(2-\alpha)} \varepsilon^{1-\alpha} n^{1/\alpha} \int_0^1 f(s)^\alpha ds + o_P(n^{1/\alpha}),$$

which, combined with (36), implies the result. \square

3.3 Proof of Theorem 2

Let $f \in \mathcal{F}$. We also assume that $f(x)$ is decreasing, $x^2 f(x)$ is increasing and $f(x) \geq 1$ for all x ; see the remark at the beginning of Section 3.

Let

$$Z_n = \int_{(0,1]} f(x) M_n(dx). \quad (38)$$

We have to show that

$$\gamma \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) = n \int_0^1 f(s) ds - n^{1/\alpha} Z_n + o_P(n^{1/\alpha}).$$

In view of Proposition 14 it suffices to show that

$$\sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) (Y_k - \gamma) = n^{1/\alpha} Z_n + o_P(n^{1/\alpha}). \quad (39)$$

Enumerate the points of Θ_n as $(x_i, y_i)_{i=1}^\infty$ and for $\varepsilon > 0$ let

$$S_n(\varepsilon) = \sum_{f(x_i) y_i > \varepsilon} f(x_i) y_i. \quad (40)$$

First let us check that

$$S_n(\varepsilon) - E(S_n(\varepsilon)) \rightarrow Z_n$$

in probability as $\varepsilon \rightarrow 0$. For this purpose note that from (6)

$$E(S_n(\varepsilon)) = \int_0^1 \int_{\varepsilon/f(x)}^\infty f(x) y \nu(dx, dy) = \varepsilon^{1-\alpha} \frac{\alpha}{\Gamma(2-\alpha)} \int_0^1 f(x)^\alpha dx, \quad (41)$$

which is finite by our assumptions on f . Thus the sum in (40) has a.s. finitely many summands. Also for $\eta > 0$

$$\text{Var}\left(\sum_{i \geq 1} f(x_i) y_i \cdot 1_{\{\eta \leq y_i, f(x_i) y_i \leq \varepsilon\}}\right) = \int_0^1 \int_\eta^{\varepsilon/f(x)} f(x)^2 y^2 \cdot 1_{\{\eta \leq \varepsilon/f(x)\}} \nu(dx, dy).$$

Letting $\eta \rightarrow 0$ we obtain

$$\text{Var}(Z_n - S_n(\varepsilon)) = \int_0^1 \int_0^{\varepsilon/f(x)} f(x)^2 y^2 \nu(dx, dy) = \varepsilon^{2-\alpha} \frac{\alpha(\alpha-1)}{\Gamma(3-\alpha)} \int_0^1 f(x)^\alpha dx,$$

which goes to 0 as $\varepsilon \rightarrow 0$.

It now follows from Chebyshev's Inequality and the fact that $E[Z] = 0$ that for all $\eta > 0$ and sufficiently small ε ,

$$P(|Z_n - (S_n(\varepsilon) - E(S_n(\varepsilon)))| > \eta) < \eta.$$

Because $E(S_n(\varepsilon))$ equals the constant d in (35), it follows from Lemma 15 that for sufficiently small ε we have

$$P\left(\left|n^{-1/\alpha} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) (Y_k 1_{\{f(X_k/n)Y_k \leq \varepsilon n^{1/\alpha}\}} - \gamma) - (Z_n - S_n(\varepsilon))\right| > 2\eta\right) \leq 2\eta$$

for sufficiently large n . Thus, to show (39), it suffices to show that for all $\varepsilon > 0$ we have

$$P\left(\left|n^{-1/\alpha} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) Y_k 1_{\{f(X_k/n)Y_k > \varepsilon n^{1/\alpha}\}} - S_n(\varepsilon)\right| > \eta\right) < \eta \quad (42)$$

for sufficiently large n . For this, we will use the following lemmas. Again let Θ_n be the Poisson point process constructed in section 2.2 from another Poisson point process Ψ_n . Denote the points of Θ_n by $(x_i, y_i)_{i=1}^\infty$ with $x_i = m(r_i)$ and $y_i = m(r_i)u_i$, where (s_i, u_i) are the points of Ψ_n and $r_i = -s_i$.

Lemma 16. *Let $\delta > 0$, $\varepsilon > 0$. With probability tending to 1 as $n \rightarrow \infty$, for all i such that $f(x_i)y_i > \varepsilon$, there exists a positive integer k_i such that the following hold:*

- (i) $R_{k_i} = n^{1-\alpha}r_i$, i.e. at time $n^{1-\alpha}r_i$ there is a merger in the coalescent back from time 0.
- (ii) The block size $X_{k_i} = N_n(n^{1-\alpha}r_i-)$ and the merger's size $Y_{k_i} = N_n(n^{1-\alpha}r_i-) - N_n(n^{1-\alpha}r_i)$ fulfil

$$(1-\delta)x_i \leq \frac{X_{k_i}}{n} \leq (1+\delta)x_i, \quad (1-\delta)y_i \leq \frac{Y_{k_i}}{n^{1/\alpha}} \leq (1+\delta)y_i.$$

Proof. The points $(n^{1-\alpha}s_i, n^{-1+1/\alpha}u_i)$ are points of Υ' . Consider those points, which in addition fulfil $f(x_i)y_i > \varepsilon$. First we verify that the probability that some of these points do not belong to Υ_n is asymptotically vanishing. The expected number of indices i with $f(x_i)y_i > \varepsilon$, that is $u_i > \varepsilon/(f(m(r_i))m(r_i))$, such that $(n^{1-\alpha}s_i, n^{-1+1/\alpha}u_i)$ is not a point of Υ_n is

$$\int_0^\infty \int_{\varepsilon/(f(m(r))m(r))}^\infty \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} u^{-1-\alpha} (1 - q(n^{-1+1/\alpha}u)) du dr.$$

Note that $1 - q(u) \leq cu$ for all $u \geq 0$. Using this bound when $r \leq n^\gamma$ and the bound $1 - q(u) \leq 1$ when $r \geq n^\gamma$, we see that the above expectation is at most

$$\begin{aligned} & cn^{-1+1/\alpha} \int_0^{n^\gamma} \int_{\varepsilon/(f(m(r))m(r))}^\infty u^{-\alpha} du dr + c \int_{n^\gamma}^\infty \int_{\varepsilon/(f(m(r))m(r))}^\infty u^{-1-\alpha} du dr \\ &= cn^{-1+1/\alpha} \int_0^{n^\gamma} \left(\frac{f(m(r))m(r)}{\varepsilon}\right)^{\alpha-1} dr + c \int_{n^\gamma}^\infty \left(\frac{f(m(r))m(r)}{\varepsilon}\right)^\alpha dr. \end{aligned}$$

Making the substitution $x = m(r)$ and using (27) again, we get that this expression is bounded above by

$$cn^{-1+1/\alpha} \int_{m(n^\gamma)}^1 x^{-1} f(x)^{\alpha-1} dx + c \int_0^{m(n^\gamma)} f(x)^\alpha dx \leq cn^{-1+1/\alpha} m(n^\gamma)^{-\zeta(\alpha-1)} + c \int_0^{m(n^\gamma)} f(x)^\alpha dx,$$

which tends to zero as $n \rightarrow \infty$ if $\gamma > 0$ is sufficiently small. By Markov's Inequality, the probability that $(n^{1-\alpha} s_i, n^{-1+1/\alpha} u_i)$ is a point of Υ_n for all i with $f(x_i) y_i > \varepsilon$ tends to 1 as $n \rightarrow \infty$.

Second we show that the probability that no more than one of the remaining $N_n(n^{1-\alpha} r_i -)$ blocks takes part in a merging event at time $n^{1-\alpha} r_i$ is asymptotically vanishing. We choose a function $h : \mathbb{N} \rightarrow \infty$ such that $\lim_{n \rightarrow \infty} h(n) = \infty$, and $h(n) = o(n^{\alpha-1})$. The expected number of indices i such that $r_i > h(n)$ and $f(x_i) y_i > \varepsilon$ is at most

$$\begin{aligned} \int_{h(n)}^\infty \int_{\varepsilon/(f(m(r))m(r))}^\infty \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} u^{-1-\alpha} du dr \\ = c \int_{h(n)}^\infty \left(\frac{f(m(r))m(r)}{\varepsilon} \right)^\alpha dr \leq c \int_0^{m(h(n))} f(x)^\alpha dx, \end{aligned}$$

which tends to zero as $n \rightarrow \infty$. Thus, we may assume $r_i \leq h(n)$ for all i with $f(x_i) y_i > \varepsilon$. In particular, because of $y_i > \varepsilon/f(x_i) \geq cx_i^\zeta = cm(r_i)^\zeta \geq cr_i^{-\zeta/(\alpha-1)} \geq ch(n)^{-\zeta/(\alpha-1)}$, this implies by our assumptions on ζ and $h(n)$ that

$$n^{1/\alpha} y_i \geq cn^{1/\alpha} h(n)^{-\zeta/(\alpha-1)} \rightarrow \infty. \quad (43)$$

Also, because $E(S_n(\varepsilon))$ does not depend on n and is finite, it suffices to show that points (x_i, y_i) with $f(x_i) y_i > \varepsilon$ and $r_i \leq h(n)$ lead to mergers fulfilling condition (ii) with high probability for large n .

At time $n^{1-\alpha} r_i$, the number of blocks of the beta coalescent is reduced by $(A_i - 1) \vee 0$, where A_i has a binomial distribution with parameters $n_i = N_n(n^{1-\alpha} r_i -)$ and $p_i = n^{-1+1/\alpha} u_i$. By Chebyshev's Inequality, we have

$$P(|A_i - n_i p_i| > \theta n^{1/\alpha} y_i \mid n_i, p_i, y_i) \leq \frac{n_i p_i (1 - p_i)}{(\theta n^{1/\alpha} y_i)^2} \leq \frac{n_i p_i}{(\theta n^{1/\alpha} y_i)^2}. \quad (44)$$

Let $\theta = \delta/3$. Lemma 13 implies that with probability tending to 1 as $n \rightarrow \infty$, we have

$$(1 - \theta) m(r_i) n \leq N_n(n^{1-\alpha} r_i -) \leq (1 + \theta) m(r_i) n, \quad (45)$$

and on this event the right-hand side of (44) is in view of (43) bounded above by

$$\frac{(1 + \theta) m(r_i) n^{1/\alpha} u_i}{(\theta n^{1/\alpha} y_i)^2} = \frac{1 + \theta}{\theta^2 n^{1/\alpha} y_i} \leq c \frac{(1 + \theta)}{\theta^2 n^{1/\alpha}} h(n)^{\frac{\zeta}{\alpha-1}} = o(n^{\zeta-1/\alpha}).$$

Due to our assumption on ζ the right-hand side tends to zero as $n \rightarrow \infty$. Taking expectations in (44) gives

$$P(|A_i - n_i p_i| > \theta n^{1/\alpha} y_i) = o(1).$$

Combining this bound with (45) gives

$$\lim_{n \rightarrow \infty} P(|A_i - n^{1/\alpha} y_i| > 2\theta n^{1/\alpha} y_i) = 0.$$

Because of (43) we get

$$\lim_{n \rightarrow \infty} P(|(A_i - 1) - n^{1/\alpha} y_i| > 3\theta n^{1/\alpha} y_i) = 0. \quad (46)$$

Thus, with probability tending to one as $n \rightarrow \infty$, the beta coalescent must have a merger at time $n^{1-\alpha} r_i$, which means $n^{1-\alpha} r_i = R_{k_i}$ for some positive integer k_i . That is, condition (i) in the statement of the lemma holds. Now, because $\delta = 3\theta$, condition (ii) is a consequence of Lemma 13 and (46). \square

Lemma 17. *Let $\delta > 0$, $\varepsilon > 0$. With probability going to 1 as $n \rightarrow \infty$, for all $0 \leq k < \tau_n$ with $f(\frac{X_k}{n}) \frac{Y_k}{n^{1/\alpha}} > \varepsilon$,*

$$(1 - \delta)x_{i_k} \leq \frac{X_k}{n} \leq (1 + \delta)x_{i_k}, \quad (1 - \delta)y_{i_k} \leq \frac{Y_k}{n^{1/\alpha}} \leq (1 + \delta)y_{i_k}, \quad (47)$$

where i_k is determined by $X_k = N_n(n^{1-\alpha} r_{i_k} -)$ or $Y_k = N_n(n^{1-\alpha} r_{i_k} -) - N_n(n^{1-\alpha} r_{i_k})$.

Proof. First we estimate the expectation of

$$S_n(\varepsilon, \gamma) = n^{-1/\alpha} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) Y_k 1_{\{f(X_k/n) Y_k > \varepsilon n^{1/\alpha}\}} 1_{\{X_k \leq \gamma n\}}$$

with $0 < \gamma \leq 1$. From Lemma 10 and (28)

$$E(Y_0 1_{\{f(X_0/n) Y_0 > \varepsilon n^{1/\alpha}\}} \mid X_0 = j) \leq c \sum_{i > \varepsilon n^{1/\alpha} / f(j/n)} i q_i \leq c \left(\frac{\varepsilon n^{1/\alpha}}{f(j/n)} \right)^{1-\alpha}.$$

Consequently

$$E\left(f\left(\frac{X_k}{n}\right) Y_k 1_{\{f(X_k/n) Y_k > \varepsilon n^{1/\alpha}\}} 1_{\{X_k \leq \gamma n\}}\right) \leq c \varepsilon^{1-\alpha} n^{1/\alpha-1} E\left(f\left(\frac{X_k}{n}\right)^\alpha 1_{\{X_k \leq \gamma n\}}\right)$$

and, since f is assumed to be decreasing,

$$\begin{aligned} E(S_n(\varepsilon, \gamma)) &\leq c \varepsilon^{1-\alpha} n^{-1} \sum_{k=0}^n E\left(f\left(\frac{X_k}{n}\right)^\alpha 1_{\{X_k \leq \gamma n\}}\right) \\ &\leq c \varepsilon^{1-\alpha} n^{-1} \sum_{1 \leq j \leq \gamma n} f\left(\frac{j}{n}\right)^\alpha \leq c \varepsilon^{1-\alpha} \int_0^\gamma f(s)^\alpha ds. \end{aligned}$$

By the assumptions on f , the integral is finite.

For $\gamma = 1$ we see that $E(S_n(\varepsilon, \gamma))$ is uniformly bounded in n . Since each positive summand of $S_n(\varepsilon, \gamma)$ is bigger than ε , it follows that the number of positive summands is stochastically bounded. Therefore it is sufficient to verify that for any $0 \leq k < \tau_n$ with $f(X_k/n) Y_k > \varepsilon n^{1/\alpha}$ the two formulas (47) hold with probability going to 1, uniformly in k .

Let $\theta > 0$. Then there is a $\gamma > 0$ such that $E(S_n(\varepsilon, \gamma)) \leq \theta \varepsilon$. Therefore, with probability at least $1 - \theta$, we have $X_k \geq \gamma n$ for all k such that $f(X_k/n) Y_k > \varepsilon n^{1/\alpha}$. In view of Lemma 13, since $N_n(r)$ is decreasing, this implies that with probability going to 1 we have $(1 + \delta)m(r_{i_k}) \geq \gamma$, which implies that $r_{i_k} \leq T$ for some fixed constant $T < \infty$ and also

$$(1 - \delta)x_{i_k} \leq \frac{X_k}{n} \leq (1 + \delta)x_{i_k}.$$

Furthermore $X_k \geq \gamma n$ and $f(X_k/n)Y_k > \varepsilon n^{1/\alpha}$ imply

$$Y_k > cn^{1/\alpha}$$

with $c = \varepsilon / \sup_{x \geq \gamma} f(x) > 0$. Let $n_k = N_n(n^{1-\alpha}r_{i_k} -)$. Since $r_{i_k} \leq T$, by Lemma 13 with probability going to 1

$$|n_k p_{i_k} - n^{1/\alpha} y_{i_k}| = |n^{-1} N_n(n^{1-\alpha} r_{i_k} -) - m(r_{i_k})| n^{1/\alpha} u_{i_k} \leq \frac{\delta}{3} n^{1/\alpha} y_{i_k}. \quad (48)$$

Thus for $\eta = 1/(2\alpha)$ and n sufficiently large

$$\begin{aligned} P(Y_k > cn^{1/\alpha}, |Y_k - n^{1/\alpha} y_{i_k}| > \delta y_{i_k} n^{1/\alpha} \mid n_k, p_{i_k}, y_{i_k}) \\ \leq P(Y_k > 0, |Y_k + 1 - n_k p_{i_k}| > \frac{\delta}{2} y_{i_k} n^{1/\alpha} \mid n_k, p_{i_k}, y_{i_k}) 1_{\{y_{i_k} > n^{-\eta}\}} \\ + P(Y_k > cn^{1/\alpha} \mid n_k, p_{i_k}, y_{i_k}) 1_{\{y_{i_k} \leq n^{-\eta}\}}. \end{aligned}$$

Since $Y_k = (A_k - 1) \vee 0$, where A_k is binomial with parameters n_k and p_{i_k} , it follows that

$$\begin{aligned} P(Y_k > cn^{1/\alpha}, |Y_k - n^{1/\alpha} y_{i_k}| > \delta y_{i_k} n^{1/\alpha} \mid n_k, p_{i_k}, y_{i_k}) \\ \leq P(Y_k > 0, |A_k - n_k p_{i_k}| > \frac{\delta}{2} y_{i_k} n^{1/\alpha} \mid n_k, p_{i_k}, y_{i_k}) 1_{\{y_{i_k} > n^{-\eta}\}} \\ + P(A_k > cn^{1/\alpha} \mid n_k, p_{i_k}, y_{i_k}) 1_{\{y_{i_k} \leq n^{-\eta}\}}. \end{aligned}$$

Chebyshev's and Markov's inequality imply together with (48)

$$\begin{aligned} P(Y_k > cn^{1/\alpha}, |Y_k - n^{1/\alpha} y_{i_k}| > \delta y_{i_k} n^{1/\alpha} \mid n_k, p_{i_k}, y_{i_k}) \\ \leq \frac{4}{\delta^2} \frac{n_k p_{i_k} (1 - p_{i_k})}{(y_{i_k} n^{1/\alpha})^2} 1_{\{y_{i_k} > n^{-\eta}\}} + \frac{n_k p_{i_k}}{cn^{1/\alpha}} 1_{\{y_{i_k} \leq n^{-\eta}\}} \\ \leq \frac{8}{\delta^2} \frac{1}{y_{i_k} n^{1/\alpha}} 1_{\{y_{i_k} > n^{-\eta}\}} + \frac{2y_{i_k}}{c} 1_{\{y_{i_k} \leq n^{-\eta}\}} \leq cn^{-1/(2\alpha)} \end{aligned}$$

and therefore

$$P(Y_k > cn^{1/\alpha}, |Y_k - n^{1/\alpha} y_{i_k}| > \delta y_{i_k} n^{1/\alpha}) \rightarrow 0.$$

Since θ was arbitrary, this gives our assertions. \square

Proof of Theorem 2. We make use of the following preliminary estimate for f . Let $\xi > 0$. Also let $0 < \delta < 1$, $u, U \in (0, 1]$, and $v, V \geq 0$. If $U \geq (1 - \delta)u$ and $V \leq (1 + \delta)v$, then, since $f(x)$ is decreasing and $x^2 f(x)$ is increasing,

$$f(U)V \leq f((1 - \delta)u)(1 + \delta)v \leq (1 - \delta)^{-2}(1 + \delta)f(u)v \leq (1 + \xi)f(u)v,$$

if δ is sufficiently small (given ξ). Together with an analogous, reversed estimate we get that if $\xi > 0$, and $(1 - \delta)u \leq U \leq (1 + \delta)u$, and $(1 - \delta)v \leq V \leq (1 + \delta)v$, then

$$(1 - \xi)f(u)v \leq f(U)V \leq (1 + \xi)f(u)v,$$

if $\delta > 0$ is sufficiently small, which entails

$$|f(U)V - f(u)v| \leq \xi f(u)v.$$

Now let $\varepsilon, \xi > 0$. Then by Lemmas 16 and 17 it follows that, if $f(x_i)y_i \notin [\varepsilon - \xi, \varepsilon + \xi]$ for all $i \geq 1$, then with probability going to 1

$$f(x_i)y_i > \varepsilon \quad \Rightarrow \quad f\left(\frac{X_{k_i}}{n}\right)\frac{Y_{k_i}}{n^{1/\alpha}} > \varepsilon \quad \text{and} \quad f\left(\frac{X_k}{n}\right)\frac{Y_k}{n^{1/\alpha}} > \varepsilon \quad \Rightarrow \quad f(x_{i_k})y_{i_k} > \varepsilon.$$

Thus, with $S_n(\varepsilon)$ defined as in (40),

$$\begin{aligned} & \left| n^{-1/\alpha} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) Y_k 1_{\{f(X_k/n)Y_k > \varepsilon n^{1/\alpha}\}} - S_n(\varepsilon) \right| \\ &= \left| \sum_{i \geq 1} \left(f\left(\frac{X_{k_i}}{n}\right) \frac{Y_{k_i}}{n^{1/\alpha}} - f(x_i)y_i \right) 1_{\{f(x_i)y_i > \varepsilon\}} \right| \\ &\leq \xi \sum_{i \geq 1} f(x_i)y_i 1_{\{f(x_i)y_i > \varepsilon\}} = \xi S_n(\varepsilon). \end{aligned}$$

We end up with

$$\begin{aligned} & P\left(\left| n^{-1/\alpha} \sum_{k < \tau_n} f\left(\frac{X_k}{n}\right) Y_k 1_{\{f(X_k/n)Y_k > \varepsilon n^{1/\alpha}\}} - S_n(\varepsilon) \right| > \eta \right) \\ &\leq P(\xi S_n(\varepsilon) > \eta) + P(\varepsilon - \xi \leq f(x_i)y_i \leq \varepsilon + \xi \text{ for some } i \geq 1) + o(1) \\ &\leq \frac{\xi}{\eta} E(S_n(\varepsilon)) + P(\varepsilon - \xi \leq f(x_i)y_i \leq \varepsilon + \xi \text{ for some } i \geq 1) + o(1). \end{aligned}$$

Since the expectation is uniformly bounded in n by (41), the first term can be made arbitrarily small by decreasing ξ . The same is true for the second term, thus (42) follows, and the proof is finished. \square

References

- [1] J. Berestycki, N. Berestycki, and J. Schweinsberg (2007). Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35**, 1835-1887.
- [2] J. Berestycki, N. Berestycki, and J. Schweinsberg (2008). Small-time behavior of beta coalescents. *Ann. Inst. H. Poincaré Probab. Statist.* **44**, 214-238.
- [3] J. Bertoin and J.-F. Le Gall (2006). Stochastic flows associated to coalescent processes III: Limit theorems. *Illinois J. Math.* **50**, 147-181.
- [4] E. Bolthausen and A.-S. Sznitman (1998). On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197**, 247-276.
- [5] I. Dahmer, G. Kersting, and A. Wakolbinger (2013). The total external branch length of beta coalescents. arXiv:1212.6070, to appear in *Comb. Probab. Comput.*,
- [6] J.-F. Delmas, J.-S. Dhersin, and A. Siri-Jegousse (2008). Asymptotic results on the length of coalescent trees. *Ann. Probab.* **18**, 997-1025.
- [7] A. Gnedin and Y. Yakubovich (2007). On the number of collisions in Λ -coalescents. *Electron. J. Probab.* **12**, 1547-1567.

- [8] A. Greven, P. Pfaffelhuber, and A. Winter (2013). Tree-valued resampling dynamics: Martingale problems and applications. *Probab. Theory Related Fields* **155**, 789-838.
- [9] G. Kersting (2012). The asymptotic distribution of the length of beta coalescent trees. *Ann. Appl. Probab.* **22**, 2086-2107.
- [10] J. F. C. Kingman (1982). The coalescent. *Stochastic Process. Appl.* **13**, 235-248.
- [11] A. Iksanov and M. Möhle (2008). On the number of jumps of random walks with a barrier. *Adv. Appl. Probab.* **40**, 206-228.
- [12] P. Pfaffelhuber and A. Wakolbinger (2006). The process of most recent common ancestors in an evolving coalescent. *Stochastic Process. Appl.* **116**, 1836-1859.
- [13] P. Pfaffelhuber, A. Wakolbinger, and H. Weisshaupt (2009). The tree length of an evolving coalescent. *Probab. Theory Related Fields* **151**, 529-557.
- [14] J. Pitman (1999). Coalescents with multiple collisions. *Ann. Probab.* **27**, 1870-1902.
- [15] S. Sagitov (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, 1116-1125.
- [16] G. Samorodnitsky and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York.
- [17] J. Schweinsberg (2003). Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.* **106**, 107-139.
- [18] J. Schweinsberg (2012). Dynamics of the evolving Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **17**, no. 91, 1-50.
- [19] M. Steinrücken, M. Birkner, and J. Blath (2013). Analysis of DNA sequence variation within marine species using Beta-coalescents. *Theor. Pop. Biol.* **87**, 15-24.